

## **RAW SEQUENCE LISTING**

**The Biotechnology Systems Branch of the Scientific and Technical  
Information Center (STIC) no errors detected.**

Application Serial Number: 09/369,735C  
Source: 1FW16  
Date Processed by STIC: 5/9/05

# ***ENTERED***

# CRF Errors Edited by the STIC Systems Branch

Serial Number: 09/369,735C

CRF Edit Date: 5/10/05  
Edited by: AZ

\_\_\_ Realigned nucleic acid/amino acid numbers/text in cases where the sequence text "wrapped" to the next line

\_\_\_ Corrected the SEQ ID NO. Sequence numbers edited were:

\_\_\_\_\_

\_\_\_ Inserted or corrected a nucleic number at the end of a nucleic line. SEQ ID NO's edited:

\_\_\_\_\_

\_\_\_ Deleted: / invalid beginning/end-of-file text ; \_\_\_ page numbers

\_\_\_ Inserted mandatory headings/numeric identifiers, specifically:

\_\_\_\_\_

\_\_\_ Moved responses to same line as heading/numeric identifier, specifically:

\_\_\_\_\_

\_\_\_ Other:

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_



IFW16

## RAW SEQUENCE LISTING

DATE: 05/10/2005

PATENT APPLICATION: US/09/369,735C

TIME: 15:11:04

Input Set : N:\AMC\369735.txt

Output Set: N:\CRF4\05102005\I369735C.raw

```

4 <110> APPLICANT: Matsui, Ikuo
5     Ishikawa, Kazuhiko
6     Ishida, Hiroyasu
7     Kosugi, Yoshitsugu
9 <120> TITLE OF INVENTION: METHODS FOR MAKING AND USING A THERMOPHILIC
10    ENZYME AS A BETA-GLYCOSIDASE (AMENDED)
12 <130> FILE REFERENCE: 11059/002001
14 <140> CURRENT APPLICATION NUMBER: 09/369,735C
15 <141> CURRENT FILING DATE: 1999-08-06
17 <160> NUMBER OF SEQ ID NOS: 9
19 <170> SOFTWARE: FastSEQ for Windows Version 4.0
21 <210> SEQ ID NO: 1
22 <211> LENGTH: 1269
23 <212> TYPE: DNA
24 <213> ORGANISM: Pyrococcus horikoshii
26 <220> FEATURE:
27 <221> NAME/KEY: CDS
28 <222> LOCATION: (1)...(1269)
30 <400> SEQUENCE: 1
31 atg ccg ctg aaa ttc ccg gaa atg ttt ctc ttt ggt acc gca aca tca      48
32 Met Pro Leu Lys Phe Pro Glu Met Phe Leu Phe Gly Thr Ala Thr Ser
33 1          5          10          15
35 tcc cat cag ata gag gga aat aat aga tgg aat gat tgg tgg tac tat      96
36 Ser His Gln Ile Glu Gly Asn Asn Arg Trp Asn Asp Trp Trp Tyr Tyr
37          20          25          30
39 gag cag att gga aag ctc ccc tac aga tct ggt aag gct tgc aat cac      144
40 Glu Gln Ile Gly Lys Leu Pro Tyr Arg Ser Gly Lys Ala Cys Asn His
41          35          40          45
43 tgg gaa ctt tac agg gat gat att cag cta atg acc agc ttg ggc tat      192
44 Trp Glu Leu Tyr Arg Asp Asp Ile Gln Leu Met Thr Ser Leu Gly Tyr
45          50          55          60
47 aat gct tat agg ttc tcc ata gag tgg agc agg cta ttc cca gag gaa      240
48 Asn Ala Tyr Arg Phe Ser Ile Glu Trp Ser Arg Leu Phe Pro Glu Glu
49 65          70          75          80
51 aat aaa ttt aat gaa gat gct ttc atg aaa tac cgg gag att ata gac      288
52 Asn Lys Phe Asn Glu Asp Ala Phe Met Lys Tyr Arg Glu Ile Ile Asp
53          85          90          95
55 ttg tta ttg acg aga ggt ata act ccc ctg gtg acc cta cac cac ttt      336
56 Leu Leu Leu Thr Arg Gly Ile Thr Pro Leu Val Thr Leu His His Phe
57          100          105          110
59 act agc cct ctc tgg ttc atg aag aaa ggt ggc ttc ctt agg gag gag      384
60 Thr Ser Pro Leu Trp Phe Met Lys Lys Gly Gly Phe Leu Arg Glu Glu
61          115          120          125

```

## RAW SEQUENCE LISTING

DATE: 05/10/2005

PATENT APPLICATION: US/09/369,735C

TIME: 15:11:04

Input Set : N:\AMC\369735.txt

Output Set: N:\CRF4\05102005\I369735C.raw

```

63 aac cta aaa cat tgg gaa aag tac ata gaa aag gtt gct gag ctt tta      432
64 Asn Leu Lys His Trp Glu Lys Tyr Ile Glu Lys Val Ala Glu Leu Leu
65      130      135      140
67 gaa aaa gtt aaa cta gta gct acc ttc aat gag ccg atg gta tac gta      480
68 Glu Lys Val Lys Leu Val Ala Thr Phe Asn Glu Pro Met Val Tyr Val
69 145      150      155      160
71 atg atg gga tat cta acg gct tat tgg ccc cca ttc att agg agt cca      528
72 Met Met Gly Tyr Leu Thr Ala Tyr Trp Pro Pro Phe Ile Arg Ser Pro
73      165      170      175
75 ttt aag gcc ttt aag gta gct gca aac ctg ctt aaa gct cac gca att      576
76 Phe Lys Ala Phe Lys Val Ala Ala Asn Leu Leu Lys Ala His Ala Ile
77      180      185      190
79 gcc tat gaa ctt ctt cat ggg aaa ttc aaa gtt gga atc gta aag aat      624
80 Ala Tyr Glu Leu Leu His Gly Lys Phe Lys Val Gly Ile Val Lys Asn
81      195      200      205
83 att ccc ata ata ctc cca gcg agt gac aag gag agg gat aga aaa gcc      672
84 Ile Pro Ile Ile Leu Pro Ala Ser Asp Lys Glu Arg Asp Arg Lys Ala
85      210      215      220
87 gct gag aaa gct gat aat tta ttt aac tgg cac ttt ttg gat gcg ata      720
88 Ala Glu Lys Ala Asp Asn Leu Phe Asn Trp His Phe Leu Asp Ala Ile
89 225      230      235      240
91 tgg agt ggg aaa tac aga ggg gta ttt aaa aca tat agg att ccc caa      768
92 Trp Ser Gly Lys Tyr Arg Gly Val Phe Lys Thr Tyr Arg Ile Pro Gln
93      245      250      255
95 agt gac gca gat ttc att ggg gtt aac tat tac acg gcc agc gaa gta      816
96 Ser Asp Ala Asp Phe Ile Gly Val Asn Tyr Tyr Thr Ala Ser Glu Val
97      260      265      270
99 agg cat act tgg aat cct tta aaa ttc ttc ttt gag gtg aaa tta gcg      864
100 Arg His Thr Trp Asn Pro Leu Lys Phe Phe Phe Glu Val Lys Leu Ala
101      275      280      285
103 gat att agc gag agg aag act caa atg gga tgg agc gtt tat cca aaa      912
104 Asp Ile Ser Glu Arg Lys Thr Gln Met Gly Trp Ser Val Tyr Pro Lys
105      290      295      300
107 gga ata tac atg gcc ctt aaa aaa gct tcc agg tat gga agg cct ctt      960
108 Gly Ile Tyr Met Ala Leu Lys Lys Ala Ser Arg Tyr Gly Arg Pro Leu
109 305      310      315      320
111 tat att acg gaa aac gga ata gcg acg ctt gat gat gaa tgg aga gtg      1008
112 Tyr Ile Thr Glu Asn Gly Ile Ala Thr Leu Asp Asp Glu Trp Arg Val
113      325      330      335
115 gaa ttc ata att caa cac ctc caa tac gtt cat aag gct atc gaa gac      1056
116 Glu Phe Ile Ile Gln His Leu Gln Tyr Val His Lys Ala Ile Glu Asp
117      340      345      350
119 ggc ctg gat gta aga ggt tac ttc tat tgg tca ttt atg gat aac tac      1104
120 Gly Leu Asp Val Arg Gly Tyr Phe Tyr Trp Ser Phe Met Asp Asn Tyr
121      355      360      365
123 gag tgg aaa gag ggg ttt ggg cct aga ttt ggc cta gtg gaa gtt gat      1152
124 Glu Trp Lys Glu Gly Phe Gly Pro Arg Phe Gly Leu Val Glu Val Asp
125      370      375      380
127 tat caa acc ttc gag aga agg ccc agg aag agt gct tac gta tac gga      1200

```

## RAW SEQUENCE LISTING

DATE: 05/10/2005

PATENT APPLICATION: US/09/369,735C

TIME: 15:11:04

Input Set : N:\AMC\369735.txt

Output Set: N:\CRF4\05102005\I369735C.raw

```

128 Tyr Gln Thr Phe Glu Arg Arg Pro Arg Lys Ser Ala Tyr Val Tyr Gly
129 385          390          395          400
131 gaa att gca aga agt aag gaa ata aag gat gag cta tta aag aga tat      1248
132 Glu Ile Ala Arg Ser Lys Glu Ile Lys Asp Glu Leu Leu Lys Arg Tyr
133          405          410          415
135 ggc cta cca gaa ctt caa ctt      1269
136 Gly Leu Pro Glu Leu Gln Leu
137          420
140 <210> SEQ ID NO: 2
141 <211> LENGTH: 423
142 <212> TYPE: PRT
143 <213> ORGANISM: Pyrococcus horikoshii
145 <400> SEQUENCE: 2
146 Met Pro Leu Lys Phe Pro Glu Met Phe Leu Phe Gly Thr Ala Thr Ser
147 1          5          10          15
148 Ser His Gln Ile Glu Gly Asn Asn Arg Trp Asn Asp Trp Trp Tyr Tyr
149          20          25          30
150 Glu Gln Ile Gly Lys Leu Pro Tyr Arg Ser Gly Lys Ala Cys Asn His
151          35          40          45
152 Trp Glu Leu Tyr Arg Asp Asp Ile Gln Leu Met Thr Ser Leu Gly Tyr
153          50          55          60
154 Asn Ala Tyr Arg Phe Ser Ile Glu Trp Ser Arg Leu Phe Pro Glu Glu
155 65          70          75          80
156 Asn Lys Phe Asn Glu Asp Ala Phe Met Lys Tyr Arg Glu Ile Ile Asp
157          85          90          95
158 Leu Leu Leu Thr Arg Gly Ile Thr Pro Leu Val Thr Leu His His Phe
159          100          105          110
160 Thr Ser Pro Leu Trp Phe Met Lys Lys Gly Gly Phe Leu Arg Glu Glu
161          115          120          125
162 Asn Leu Lys His Trp Glu Lys Tyr Ile Glu Lys Val Ala Glu Leu Leu
163          130          135          140
164 Glu Lys Val Lys Leu Val Ala Thr Phe Asn Glu Pro Met Val Tyr Val
165 145          150          155          160
166 Met Met Gly Tyr Leu Thr Ala Tyr Trp Pro Pro Phe Ile Arg Ser Pro
167          165          170          175
168 Phe Lys Ala Phe Lys Val Ala Ala Asn Leu Leu Lys Ala His Ala Ile
169          180          185          190
170 Ala Tyr Glu Leu Leu His Gly Lys Phe Lys Val Gly Ile Val Lys Asn
171          195          200          205
172 Ile Pro Ile Ile Leu Pro Ala Ser Asp Lys Glu Arg Asp Arg Lys Ala
173          210          215          220
174 Ala Glu Lys Ala Asp Asn Leu Phe Asn Trp His Phe Leu Asp Ala Ile
175 225          230          235          240
176 Trp Ser Gly Lys Tyr Arg Gly Val Phe Lys Thr Tyr Arg Ile Pro Gln
177          245          250          255
178 Ser Asp Ala Asp Phe Ile Gly Val Asn Tyr Tyr Thr Ala Ser Glu Val
179          260          265          270
180 Arg His Thr Trp Asn Pro Leu Lys Phe Phe Phe Glu Val Lys Leu Ala
181          275          280          285

```

## RAW SEQUENCE LISTING

DATE: 05/10/2005

PATENT APPLICATION: US/09/369,735C

TIME: 15:11:04

Input Set : N:\AMC\369735.txt

Output Set: N:\CRF4\05102005\I369735C.raw

```

182 Asp Ile Ser Glu Arg Lys Thr Gln Met Gly Trp Ser Val Tyr Pro Lys
183      290      295      300
184 Gly Ile Tyr Met Ala Leu Lys Lys Ala Ser Arg Tyr Gly Arg Pro Leu
185 305      310      315      320
186 Tyr Ile Thr Glu Asn Gly Ile Ala Thr Leu Asp Asp Glu Trp Arg Val
187      325      330      335
188 Glu Phe Ile Ile Gln His Leu Gln Tyr Val His Lys Ala Ile Glu Asp
189      340      345      350
190 Gly Leu Asp Val Arg Gly Tyr Phe Tyr Trp Ser Phe Met Asp Asn Tyr
191      355      360      365
192 Glu Trp Lys Glu Gly Phe Gly Pro Arg Phe Gly Leu Val Glu Val Asp
193      370      375      380
194 Tyr Gln Thr Phe Glu Arg Arg Pro Arg Lys Ser Ala Tyr Val Tyr Gly
195 385      390      395      400
196 Glu Ile Ala Arg Ser Lys Glu Ile Lys Asp Glu Leu Leu Lys Arg Tyr
197      405      410      415
198 Gly Leu Pro Glu Leu Gln Leu
199      420
201 <210> SEQ ID NO: 3
202 <211> LENGTH: 57
203 <212> TYPE: DNA
204 <213> ORGANISM: Artificial Sequence
206 <220> FEATURE:
207 <223> OTHER INFORMATION: An upper primer designed to create the NdeI site.
209 <400> SEQUENCE: 3
210 taagaaggag atatacatat gccgctgaaa ttcccgga aa tgtttctctt tggtag      57
212 <210> SEQ ID NO: 4
213 <211> LENGTH: 46
214 <212> TYPE: DNA
215 <213> ORGANISM: Artificial Sequence
217 <220> FEATURE:
218 <223> OTHER INFORMATION: A lower primer designed to create the BamHI site.
220 <400> SEQUENCE: 4
221 ttactgcag agaggatccc taatcctaaa gttgaagttc tggtag      46
223 <210> SEQ ID NO: 5
224 <211> LENGTH: 483
225 <212> TYPE: PRT
226 <213> ORGANISM: Pyrococcus horikoshii
228 <400> SEQUENCE: 5
229 Met Lys Phe Tyr Trp Gly Val Val Gln Ser Ala Phe Gln Phe Glu Met
230 1      5      10      15
231 Gly Asp Pro Tyr Arg Arg Asn Ile Asp Pro Arg Ser Asp Trp Trp Tyr
232      20      25      30
233 Trp Val Arg Asp Pro Tyr Asn Ile Lys Asn Asp Leu Val Ser Gly Asp
234      35      40      45
235 Leu Pro Glu Glu Gly Ile Asn Asn Tyr Glu Leu Tyr Glu Ile Asp His
236      50      55      60
237 Arg Leu Ala Lys Glu Leu Gly Leu Asn Ala Tyr Gln Leu Thr Ile Glu
238 65      70      75      80

```

## RAW SEQUENCE LISTING

DATE: 05/10/2005

PATENT APPLICATION: US/09/369,735C

TIME: 15:11:04

Input Set : N:\AMC\369735.txt

Output Set: N:\CRF4\05102005\I369735C.raw

```

239 Trp Ser Arg Ile Phe Pro Cys Pro Thr Phe Asn Val Glu Val Glu Phe
240      85      90      95
241 Glu Arg Asp Asx Tyr Gly Leu Ile Lys Lys Val Lys Ile Glu Lys Glu
242      100      105      110
243 His Leu Glu Glu Leu Asp Lys Leu Ala Asn Gln Lys Glu Val Arg His
244      115      120      125
245 Tyr Leu Asn Val Leu Arg Asn Leu Lys Lys Leu Gly Phe Thr Thr Phe
246      130      135      140
247 Val Thr Leu Asn His Gln Thr Asn Pro Ile Trp Ile His Asp Pro Ile
248 145      150      155      160
249 Glu Thr Arg Gly Asn Phe Gln Lys Ala Arg Ala Pro Gly Trp Val Asp
250      165      170      175
251 Glu Arg Thr Ile Ile Glu Phe Ala Lys Tyr Ala Ala Tyr Val Ala Trp
252      180      185      190
253 Lys Phe Asp Asn Tyr Val Asp Tyr Trp Ser Thr Phe Asp Glu Pro Met
254      195      200      205
255 Val Thr Ala Glu Leu Gly Tyr Leu Ala Pro Tyr Val Gly Trp Pro Pro
256      210      215      220
257 Gly Ile Leu Asn Pro Ser Ala Ala Lys Lys Val Ile Ile Asn Gln Ile
258 225      230      235      240
259 Val Ala His Ala Pro Ala Tyr Asp Ser Ile Lys Lys Phe Ser Ser Lys
260      245      250      255
261 Pro Val Gly Val Ile Leu Asn Ile Ile Pro Ala Tyr Pro Leu Asp Pro
262      260      265      270
263 Asn Asp Ser Lys Ser Val Arg Ala Ala Glu Asn Tyr Asp Leu Phe His
264      275      280      285
265 Asn Arg Leu Phe Leu Glu Ala Val Asn Arg Gly Asn Val Asp Leu Asp
266      290      295      300
267 Ile Thr Gly Glu Tyr Thr Lys Ile Pro His Ile Lys Arg Asn Asp Trp
268 305      310      315      320
269 Ile Gly Asn Asn Tyr Tyr Thr Arg Glu Val Val Lys Tyr Val Glu Pro
270      325      330      335
271 Lys Tyr Glu Glu Leu Pro Leu Ile Thr Phe Val Gly Val Glu Gly Tyr
272      340      345      350
273 Gly Tyr Ser Gly Asn Pro Asn Ser Leu Ser Pro Asp Asn Asn Pro Thr
274      355      360      365
275 Ser Asp Phe Gly Trp Glu Val Phe Pro Gln Gly Leu Tyr Asp Ser Thr
276      370      375      380
277 Leu Glu Ala Ala Glu Tyr Asn Lys Glu Val Phe Ile Thr Glu Asn Gly
278 385      390      395      400
279 Ile Ala Asp Ser Lys Asp Ile Leu Arg Pro Arg Tyr Ile Ile Asp His
280      405      410      415
281 Val Asn Glu Val Lys Lys Leu Ile Glu Asn Gly Ile Lys Val Gly Gly
282      420      425      430
283 Tyr Phe His Trp Ala Leu Thr Asp Asn Tyr Glu Trp Ala Met Gly Phe
284      435      440      445
285 Lys Ile Arg Phe Gly Leu Tyr Glu Val Asp Leu Ile Thr Lys Glu Arg
286      450      455      460
287 Ile Pro Arg Arg Arg Ser Val Glu Ile Tyr Lys Lys Ile Val Met Glu

```

**VERIFICATION SUMMARY**

DATE: 05/10/2005

PATENT APPLICATION: US/09/369,735C

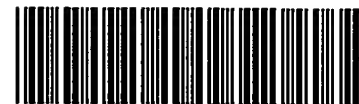
TIME: 15:11:05

Input Set : N:\AMC\369735.txt

Output Set: N:\CRF4\05102005\I369735C.raw



Raw Sequence Listing before editing,  
for reference only



IFW16

## RAW SEQUENCE LISTING

DATE: 05/09/2005

PATENT APPLICATION: US/09/369,735C

TIME: 16:52:44

Input Set : D:\11059-002001.SEQ.txt

Output Set: N:\CRF4\05092005\I369735C.raw

4 <110> APPLICANT: Matsui, Ikuo  
 5 Ishikawa, Kazuhiko  
 6 Ishida, Hiroyasu  
 7 Kosugi, Yoshitsugu  
 9 <120> TITLE OF INVENTION: METHODS FOR MAKING AND USING A THERMOPHILIC  
 10 ENZYME AS A BETA-GLYCOSIDASE (AMENDED)  
 12 <130> FILE REFERENCE: 11059/002001  
 14 <140> CURRENT APPLICATION NUMBER: 09/369,735C  
 15 <141> CURRENT FILING DATE: 1999-08-06  
 17 <160> NUMBER OF SEQ ID NOS: 9  
 19 <170> SOFTWARE: FastSEQ for Windows Version 4.0

## ERRORED SEQUENCES

499 <210> SEQ ID NO: 9  
 500 <211> LENGTH: 285  
 501 <212> TYPE: PRT  
 502 <213> ORGANISM: Artificial Sequence  
 504 <220> FEATURE:  
 505 <223> OTHER INFORMATION: Consensus  
 507 <400> SEQUENCE: 9  
 508 Met Phe Pro Glu Phe Gly Gln Ser Gly Phe Gln Phe Glu Met Gly Pro  
 509 1 5 10 15  
 510 Asp Asp Trp Trp Trp Val Asp Asn Ile Gly Leu Val Ser Gly Asp Leu  
 511 20 25 30  
 512 Pro Glu Gly Asn Trp Glu Leu Tyr Asp His Ala Lys Leu Gly Leu Asn  
 513 35 40 45  
 514 Ala Tyr Arg Ile Glu Trp Ser Arg Ile Phe Pro Pro Thr Val Glu Ile  
 515 50 55 60  
 516 Val Glu Leu Glu Leu Ala Asn Ala His Tyr Arg Ile Leu Lys Pro Gly  
 517 65 70 75 80  
 518 Thr Ile Val Asn Leu His Thr Leu Pro Asp Trp His Asp Pro Ile Arg  
 519 85 90 95  
 520 Gly Trp Leu Glu Arg Thr Val Glu Phe Ala Lys Tyr Ala Ala Tyr Val  
 521 100 105 110  
 522 Ala Lys Phe Asp Asp Val Asp Trp Ser Thr Phe Asn Glu Pro Met Val  
 523 115 120 125  
 524 Val Leu Gly Tyr Leu Tyr Ser Gly Phe Pro Pro Gly Leu Ser Pro Glu  
 525 130 135 140  
 526 Ala Ala Lys Asn Ile Ala His Ala Ala Tyr Asp Ile Lys Ser Lys Pro  
 527 145 150 155 160  
 528 Val Gly Ile Ile Tyr Asn Asp Pro Lys Asp Ala Ala Glu Phe Glu Ala

*Does Not Comply  
 Corrected Diskette Needed*

*P.2*

## RAW SEQUENCE LISTING

DATE: 05/09/2005

PATENT APPLICATION: US/09/369,735C

TIME: 16:52:44

Input Set : D:\11059-002001.SEQ.txt

Output Set: N:\CRF4\05092005\I369735C.raw

529				165				170					175			
530	Ile	Gly	Glu	Pro	Asp	Trp	Ile	Gly	Met	Asn	Tyr	Tyr	Thr	Arg	Val	Val
531				180				185					190			
532	Glu	Leu	Pro	Gly	Tyr	Gly	Leu	Ser	Pro	Ser	Asp	Phe	Gly	Trp	Glu	Tyr
533			195					200					205			
534	Arg	Glu	Gly	Leu	Tyr	Asp	Leu	Ala	Tyr	Pro	Tyr	Ile	Thr	Glu	Asn	Gly
535		210					215					220				
536	Thr	Ala	Asp	Asp	Pro	Pro	Tyr	Ile	Ser	His	Val	Lys	Ala	Ile	Glu	Gly
537	225					230					235	.				240
538	Asp	Val	Pro	Gly	Tyr	Phe	His	Trp	Ser	Leu	Thr	Asp	Asn	Tyr	Glu	Trp
539				245						250					255	
540	Ala	Gly	Glu	Met	Arg	Glu	Gly	Leu	Glu	Val	Asp	Thr	Lys	Glu	Arg	Pro
541				260					265					270		
542	Arg	Ser	Ala	Val	Tyr	Arg	Glu	Ile	Ala	Ile	Glu	Leu	Arg			
543			275					280						285		

E--> 545 9  
E--> 548 1

**VERIFICATION SUMMARY**

DATE: 05/09/2005

PATENT APPLICATION: US/09/369,735C

TIME: 16:52:45

Input Set : D:\11059-002001.SEQ.txt

Output Set: N:\CRF4\05092005\I369735C.raw

L:545 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:9

M:332 Repeated in SeqNo=9